

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年10月 9日
Date of Application:

出願番号 特願2003-350818
Application Number:
[ST. 10/C]: [JP2003-350818]

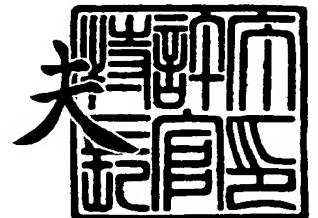
出願人 株式会社日立製作所
Applicant(s):

U.S. Appln. Filed 3-12-04
Inventor: M. Hiramatsu et al
Mattingly Stanger & Malor
Docket ASA1174

2004年 3月 2日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2004-3015585



【書類名】 特許願
【整理番号】 KN1575
【提出日】 平成15年10月 9日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 9/46
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社 日立製作所 システム開発研究所内
 【氏名】 平松 雅巳
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社 日立製作所 システム開発研究所内
 【氏名】 大島 訓
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社 日立製作所 システム開発研究所内
 【氏名】 木村 信二
【発明者】
 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社 日立製作所 ソフトウェア事業部内
 【氏名】 高杉 昌督
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社 日立製作所
【代理人】
 【識別番号】 100093492
 【弁理士】
 【氏名又は名称】 鈴木 市郎
 【電話番号】 03-3591-8550
【選任した代理人】
 【識別番号】 100078134
 【弁理士】
 【氏名又は名称】 武 顕次郎
【手数料の表示】
 【予納台帳番号】 113584
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

第1のOSと、該第1のOS上で動作し通常の業務処理を行うサービスアプリケーションと、前記第1のOSとは異なる第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記第1のOSは、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーションは、前記第1のOSが保持する情報の内容を解析して障害の兆候を検知することを特徴とする計算機システム。

【請求項 2】

前記第1のOSは、解析予測アプリケーションの補助を行う補助プログラムを有し、前記解析予測アプリケーションは、解析すべきメモリの位置と解析方法、及び、障害に対して行うべき処理の一覧を保持し、前記補助プログラムは、前記処理の一覧の内容に従い、障害の兆候によって、第1のOSの障害に対する処理を行うことを特徴とする請求項1記載の計算機システム。

【請求項 3】

前記解析予測アプリケーションは、外部端末に対して解析した障害の兆候の内容を通知することを特徴とする請求項1記載の計算機システム。

【請求項 4】

第1のOSと、該第1のOS上で動作し通常の業務処理を行うサービスアプリケーションと、前記第1のOSとは異なる第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記第1のOSが、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーションが、前記第1のOSが保持する情報の内容を解析して障害の兆候を検知する構成の計算機を複数台備え、1つの計算機の解析予測アプリケーションが、自己回復不能な障害の兆候を検知した場合、他の計算機に解析した障害の内容を通知し、処理を受け継がせることを特徴とする計算機システム。

【請求項 5】

複数の第1のOSと、該複数の第1のOSのそれぞれの上で動作し通常の業務処理を行う複数のサービスアプリケーションと、前記第1のOSとは異なる1つの第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記複数の第1のOSのそれぞれは、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーションは、前記複数第1のOSのそれぞれが保持する情報の内容を解析して障害の兆候を検知することを特徴とする計算機システム。

【請求項 6】

第1のOSと、該第1のOS上で動作し通常の業務処理を行うサービスアプリケーションと、前記第1のOSとは異なる第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記第1のOSが、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーションが、前記第1のOSが保持する情報の内容を解析して障害の兆候を検知する構成の計算機を、論理分割により仮想マルチOSセットとして複数台構成し、前記複数の仮想マルチOSセットのそれぞれの第1のOSと第2のOSとが交互に切り替えられて動作を実行し、前記複数の仮想マルチOSセットそれぞれの解析予測アプリケーションが、自セットの前記第1のOSが保持する情報の内容を解析して障害の兆候を検知することを特徴とする計算機システム。

【請求項 7】

前記複数の仮想マルチOSセットの1つの解析予測アプリケーションが、自セットの前記第1のOSでの障害の兆候を検知した場合、他の仮想マルチOSセットに解析した障害の内容を通知し、処理を受け継がせることを特徴とする請求項6記載の計算機システム。

【請求項 8】

複数の第1のOSと、該複数の第1のOSのそれぞれの上で動作し通常の業務処理を行う複数のサービスアプリケーションと、前記第1のOSとは異なる1つの第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記複数の第1のOSのそれぞれが、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーション

ンが、前記複数第 1 の O S のそれぞれが保持する情報の内容を解析して障害の兆候を検知する構成の計算機を、論理分割により仮想マルチ O S セットとして構成し、前記マルチ O S セットの解析予測アプリケーションが、前記第 1 の O S での障害の兆候を検知した場合、他方の第 1 の O S に処理を受け継ぐことを特徴とする計算機システム。

【請求項 9】

第 1 の O S とは異なる第 2 の O S 上で動作する解析予測アプリケーションが、第 1 の O S が保持している状態情報と動作記録情報との内容を解析して障害の兆候を検知することを特徴とする計算機システムの障害兆候の検知方法。

【請求項 1 0】

前記解析予測アプリケーションは、解析すべきメモリの位置と解析方法、及び、障害に対して行うべき処理の一覧を保持し、前記第 1 の O S が有する解析予測アプリケーションの補助を行う補助プログラムを使用して、前記処理の一覧の内容に従い、障害の兆候によって、第 1 の O S の障害に対する処理を行うことを特徴とする請求項 9 記載の計算機システムの障害兆候の検知方法。

【書類名】 明細書**【発明の名称】 計算機システム及び計算機システムの障害兆候の検知方法****【技術分野】****【0001】**

本発明は、計算機システム及び計算機システムの障害兆候の検知方法に係り、特に、アプリケーション（ＡＰ）、オペレーティングシステム（ＯＳ）及びハードウェア（ＨＷ）の障害発生の際を自システム内で検知可能とした計算機システム及び計算機システムの障害兆候の検知方法に関する。

【背景技術】**【0002】**

一般に、計算機システムのＡＰやＯＳは、ＡＰ、ＯＳ自身に含まれる欠陥や、ＯＳが使用している機器の障害等が主な原因となる等の様々な理由で障害を起こして停止することがある。

【0003】

前述したような障害が起こったときにも、ＡＰが提供する機能を継続する必要がある場合にＡＰの障害を検知する技術に関する従来技術として、ウォッチドッグタイマを利用して、処理終了までの時間を監視したり、通信の授受が時間内に完了したか否かを判断したりするハートビートと呼ばれる技術が知られている。また、他の従来技術として、システムが定期的に吐き出すログを監視し、障害があったことを検知するという技術が知られている。なお、一般的なＨＡクラスタで利用されているハートビートに関する従来技術として、例えば、非特許文献１に開示された技術が知られている。

【非特許文献１】 <http://www.atmarkit.co.jp/flinux/rensai/ciuster02/ciuster02.html>

【発明の開示】**【発明が解決しようとする課題】****【0004】**

前述したハートビートを利用する障害監視の方法に関する従来技術としては様々なものがあるが、何れにしても、これらの方法は、障害が実際に起こった後で、ある程度の時間が経過してから障害が検知されるため、常に障害の検知が遅くなるという問題点を有している。

【0005】

また、ハートビートやウォッチドッグタイマを用いる従来技術は、ＡＰやＯＳの負荷が高くなった場合に、処理終了までの時間、通信データの授受の時間が監視制限時間を超えてしまう場合があり、障害を誤検出するという問題点を有している。

【0006】

また、システムのログを監視する方法も、前述と同様に、障害が実際に起こった後で、ある程度の時間が経過してから障害が検知されるため、常に障害の検知が遅くなるという問題点を有している。

【0007】

さらに、前述した従来技術による障害の監視は、障害監視用のＡＰにより行われるが、障害監視のためのＡＰが監視対象のＡＰと同一のＯＳ上で実行されているため、ＯＳ自身に障害が起きた場合に、監視を行うＡＰもＯＳの障害に巻き込まれてしまい、監視機能が失われてしまうという問題点を有している。

【0008】

前述したような問題点を解決するために、監視用として別の計算機を用意するという方法もあるが、この場合も、やはり障害が起きた後でしか、障害が起きたことの検知を行うことができないという問題点が残ることになる。

【0009】

本発明の目的は、前述した従来技術の問題点を解決し、動作中のＯＳ及びＡＰの状態を解析し、障害につながる兆候を検出して、起こりうる障害に対する処理を行うことを可能

とした計算機システム及び計算機システムの障害兆候の検知方法を提供することにある。

【課題を解決するための手段】

【0010】

本発明によれば前記目的は、第1のOSと、該第1のOS上で動作し通常の業務処理を行うサービスアプリケーションと、前記第1のOSとは異なる第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記第1のOSが、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーションが、前記第1のOSが保持する情報の内容を解析して障害の兆候を検知するように計算機システムを構成することにより達成される。

【0011】

また、前記目的は、第1のOSと、該第1のOS上で動作し通常の業務処理を行うサービスアプリケーションと、前記第1のOSとは異なる第2のOSと、該第2のOS上で動作する解析予測アプリケーションとを備え、前記第1のOSが、自OSの状態情報と動作記録情報とを保持し、前記解析予測アプリケーションが、前記第1のOSが保持する情報の内容を解析して障害の兆候を検知する構成の計算機を複数台備え、1つの計算機の解析予測アプリケーションが、自己回復不能な障害の兆候を検知した場合、他の計算機に解析した障害の内容を通知し、処理を受け継がせるようにしたことにより達成される。

【発明の効果】

【0012】

本発明によれば、同時に複数のOSを動作させ、一方のOSのAPから動作中の他方のOSやAPの状態を随時解析し、障害につながる兆候を検出して、起こりうる障害に対する処理を行うようにしているので、OSやAPが不安定になり、動作不可能になったとしても、障害が発生する前にその兆候を検知し、障害に対処する処理を実行でき、障害による影響を最小限に抑えることが可能となる。

【発明を実施するための最良の形態】

【0013】

以下、本発明による計算機システムの実施形態を図面により詳細に説明する。

【0014】

図1は本発明の第1の実施形態による計算機システムのハードウェア構成を示すブロック図、図2はメモリ21の内部に格納されるプログラムの構成を示す図である。

【0015】

本発明の第1の実施形態による計算機システムは、図1に示すように、計算機20と、この計算機20に接続される外部記憶装置23、キーボード24、マウス25及び画面出力装置29とにより構成され、計算機20は、CPU22と、主記憶装置（メモリ）21と、冷却装置26と、温度センサ27と、通信装置28とを備えて構成される。

【0016】

計算機20のメモリ21の内部には、図2に示すように、CPU22により実行される第1OS1、第2OS2、サービスAP3、解析予測AP4、マルチOS制御部5の各プログラムが格納されている。そして、第1OS1は、動作記録制御部6及び補助ドライバ7を備え、マルチOS制御部5は、メモリ取得部8及びOS間メッセージ通信部9を備えている。また、第2OS2は、マルチOS制御部インタフェース10を備えている。

【0017】

図3はメモリ21に格納される各プログラム内に設けられるデータテーブルの構成を示す図である。図3に示すように、サービスAP3は、AP状態情報17を含み、第1OS1は、OS状態情報11と動作記録情報12とページ変換テーブル18とを含む。解析予測AP4は、障害兆候表13と、対処プログラム表14と、情報アドレス変換表15と、前状態情報保持部16とを含む。

【0018】

前述において、第1OS1は、一般的なOSであり、サービスAP3は、第1OS1の上で実行されるAPであり、通常の業務処理を行うプログラムである。解析予測AP4は

、第2OS2の上で実行されるAPであり、第1OS1及びサービスAP3の状態を解析して障害の兆候を検知するためのプログラムである。第2OS2は、第1OS1とは異なるOSであり、第1OS1より信頼性の高いOSであることが望ましい。

【0019】

第1OS1内の動作記録制御部6は、第1OS1が行った動作を、動作した時間と共に動作記録情報12に記録する。ここで記録される動作には、例えば、プロセスのコンテキストスイッチや、割り込み処理、システムコールやメモリ割り当て、その他第1OS1や計算機20が出力する警告やエラー動作がある。第1OS1内の補助ドライバ7は、解析予測AP4の補助を行う補助プログラムであり、解析予測AP4上に解析すべき情報が格納されるメモリの位置と解析方法及び障害に対して行うべき処理の一覧を保持し、この処理の一覧の内容に従って、障害の兆候によって、第1OS1の障害に対する処理を行う。第1OS1の外部からの処理を受け付け、実行する処理、例えば、第2OS2から要求される処理を実行する。この補助ドライバ7への指示は、マルチOS制御部5のOS間メッセージ通信部9を用いて行われる。

【0020】

マルチOS制御部インタフェース10は、第2OS2の機能として備えられるものであり、第2OS2の上で動作するAPからマルチOS制御部5を操作するインタフェースである。

【0021】

OS状態情報11は、第1OS1の現在の動作状態を表す情報である。ここに格納される情報には、第1OS1で動作するAPを管理するためのプロセス管理情報や、OSとプロセスのCPU時間の消費情報、同期を取るためのロックの使用状況、システムで開いているファイルやファイルハンドルの情報、第1OS1が管理しているメモリに関する情報、冷却装置26のファンの回転数、温度センサ27からの情報等がある。

【0022】

AP状態情報17は、サービスAP3の動作状態を表す情報であり、この情報をチェックポイントとして保存しておくことにより、サービスAP3のサービスが中断した場合に、中断した場所からサービスを再開できるだけの情報を有しているものとする。ページ変換テーブル18は、第1OS1がCPUの仮想メモリ機能を利用する場合に使用する論理アドレスから物理アドレスへの変換を行う際に参照するテーブルである。

【0023】

図6は障害兆候表13の構成を示す図である。障害兆候表13は、解析予測AP4が第1OS1やサービスAP3の情報を解析する際に使用する情報を格納しており、障害の兆候として予想される障害の兆候パターンと、その障害の兆候に対して利用する対処プログラム番号のリストとの組に兆候番号を付して保持している。例えば、兆候番号1の障害の兆候パターンが「動作記録が中断」であり、これに対処する対処プログラムリストに、プログラム番号1、4が1つの組として保持されており、また、兆候番号2の障害の兆候パターンが「デバイスからの異常発生」であり、これに対処する対処プログラムリストに、プログラム番号2、3が1つの組として保持されている。

【0024】

図7は対処プログラム表14の構成を示す図である。対処プログラム表14は、図6の障害兆候表13における対処プログラム番号のリストに保持される各対処プログラムの対処プログラム番号と、そのプログラムが実行する対処動作の内容とを保持している。

【0025】

図8は情報アドレス変換表15の構成を示す図である。情報アドレス変換表15は、第1OS1上のサービスAPが持つAP状態情報17、第1OS1が持つOS状態情報11、動作記録情報12へ第2OS2からアクセスするために必要な論理アドレスと、物理アドレスとの変換を行うための表であり、AP状態情報17、OS状態情報11、動作記録情報12を示すカーネルページの情報と、論理アドレスと、物理アドレスとの組を保持している。

【0026】

前状態情報保持部16は、図示していないが、解析予測AP4が今まで解析を行った結果や、AP状態情報17のチェックポイント情報、OS状態情報11の内、障害の兆候を調べるために必要なデータを保持している。

【0027】

マルチOS制御部5は、この上で動作する第1OS1と第2OS2とがお互いに独立して各種の処理を実行することが可能なように制御を行っている。独立して実行するとは、2つのOSが計算機の資源であるメモリ21や入出力デバイスを分割して利用し、互いの実行が他に影響を与えないことを言う。1つの計算機で複数のOSを独立に実行する技術としては、例えば、特開平11-149385号公報（以下、文献1という）に開示されている。この文献1によれば、第1OS1と第2OS2とを独立して実行でき、第1OS1が障害で停止した場合でも、第2OS2は継続して動作することができる。

【0028】

マルチOS制御部5は、さらに、あるOSから他のOSが使用している物理メモリにアクセスできるようにするメモリ取得部8を備えている。このメモリ取得部8は、例えば、あるOSの物理アドレスにあるページが、アクセス要求のあったOSからアクセス可能かどうかを調べ、要求のあったOSが利用できるように、ページテーブルに直接そのページの割り当てを行ったり、OS間でページ内容をコピーする機能である。

【0029】

また、マルチOS制御部5は、さらに、OS相互間での情報のやり取りを行うことができるように、相手のOSの手続きを呼び出す機能であるOS間メッセージ通信部9を備えている。

【0030】

図4は第2OS2が情報アドレス変換表15を作成、更新する際の処理動作を説明するフローチャートであり、次に、これについて説明する。ここでの処理は、計算機システムが立ち上げられたときに、第2OS2が第1OS1の側にあるAP状態情報17、OS状態情報11、動作記録情報12の格納されているメモリアドレスの情報を持っていないために、それらの情報が格納されているメモリアドレスを取り込んで図8に示すような情報アドレス変換表15を作成する処理である。

【0031】

(1) まず、第1OS1の側にあるAP状態情報17、OS状態情報11、動作記録情報12の格納されているメモリの論理アドレスが既に取得されて判っているか否かを判定する（ステップ101）。

【0032】

(2) ステップ101の判定で、第1OS1に格納されている動作記録情報12及びOS状態情報11、AP状態情報17の論理アドレスが取得できてなかった場合、補助ドライバ7を呼び出して動作記録情報12及びOS状態情報11、AP状態情報17の論理アドレスを取得する（ステップ102）。

【0033】

(3) ステップ101の判定で、動作記録情報12及びOS状態情報11、AP状態情報17の論理アドレスが判っていた場合、または、ステップ102の処理でこれら論理アドレスを取得した後、第1OS1がページ変換テーブル18を使用しているか否かを判定する（ステップ103）。

【0034】

(4) ステップ103の判定で、第1OS1がページ変換テーブル18を使用していた場合、このページ変換テーブル18の位置が予め判っているか否かを判定し、判らない場合、補助ドライバ7を呼び出して前述のページ変換テーブル18の位置を取得する（ステップ104、105）。

【0035】

(5) ステップ104の判定で、ページ変換テーブル18の位置が予め判っていた場合、

また、ステップ105の処理でページ変換テーブル18の位置を取得した後、メモリ取得部8によりページ変換テーブル18そのものを取得する（ステップ106）。

【0036】

(6) その後、取得したページ変換テーブル18から動作記録情報12及びOS状態情報11、AP状態情報17の論理アドレスを物理アドレスに変換し、情報アドレス変換表15を作成する（ステップ107、109）。

【0037】

(7) ステップ103の判定で、第1OS1がページ変換テーブル18を使用していなかった場合、論理アドレスを物理アドレスとして使用することとして情報アドレス変換表15を作成する（ステップ108、110）。

【0038】

前述したステップ104の判定で、ページ変換テーブル18の位置が判っていなかった場合、ページ変換テーブル18を、OSを切り替えるためにマルチOS制御部5が動作するときに、CPU22の制御レジスタから取得するようにすることもできる。このようにすることにより、ページ変換テーブル18を探し出す必要がなくなる。

【0039】

また、前述した動作記録情報12及びOS状態情報11、AP状態情報17の論理アドレス及び物理アドレスの取得方法として、第1OS1または第1OS1に内蔵された補助ドライバ7が、これらの情報にアクセスするための手段を用意し、また、これらの情報がどこにあるのかという情報を、マルチOS制御部5に対して登録するためのインタフェースをマルチOS制御部5に備えておくようにする方法もある。このようにすることにより、OS状態情報11等が動的にメモリ上の位置が変化する場合であっても対応することが可能となる。

【0040】

図5は第2OS2上の解析予測AP4の処理動作を説明するフローチャートであり、次に、これについて説明する。なお、図5に示すフローにおいて、何も記述していないステップ130は、ステップ122の繰り返しの処理を受けてステップ123の処理に戻すための処理ステップであり、また、同様に、何も記述していないステップ129は、ステップ125の繰り返しの処理を受けてステップ126の処理に戻すための処理ステップである。

【0041】

(1) まず、解析予測AP4は、マルチOS制御部5のメモリ取得部8と情報アドレス変換表15とを用いて、第1OS1のOS状態情報11や動作記録情報12、サービスAP3のAP状態情報17の内容を取得する。また、前状態情報保持部16から以前の第1OS1、サービスAP3の状態情報を取得する（ステップ121）。

【0042】

(2) 次に、ステップ121の処理で取得した情報について、障害兆候表13のパターンに従った解析を行い、ステップ121の処理で取得した情報から障害につながる動作や状態の変化のパターンを調べる。この調べは、障害兆候表13のパターンの数だけ繰り返される（ステップ122、123、130）。

【0043】

(3) 解析の結果、障害の兆候を検出したか否かを判定し、障害の兆候を検出した場合、障害兆候表13のパターンに併記された障害対処を行うプログラムリストを得て、次に説明する処理動作を繰り返す（ステップ124、125）。

【0044】

(4) まず、対処プログラム表14から対処動作プログラムを得て、この対処プログラムの行う対処動作が実行可能か否かを判定し、可能であれば、対処動作を実行する（ステップ126～128）。

【0045】

(5) ステップ127の判定で、対処動作を実行できないと判定された場合、ステップ1

26の処理に戻って、新たな対処動作を試みる（ステップ129）。

【0046】

(6) 障害兆候表にあるパターンの全てを調べ、対処動作を終了したとき、また、ステップ123、124の処理の繰り返しで、障害の兆候を検出できなかった場合、解析予測AP4は、サービスAP3を再開できるように、ステップ121の処理で取得した情報を使用して前状態情報保持部16の情報を更新し（ステップ131）、一定時間動作を停止した後、ステップ121からの処理に戻って、障害解析の処理を続ける（ステップ131、132）。

【0047】

前述した本発明の実施形態による各処理は、処理プログラムとして構成することができ、この処理プログラムは、HD、DAT、FD、MO、DVD-ROM、CD-ROM等の記録媒体に格納して提供することができ、また、ネットワークにより提供することができる。

【0048】

障害兆候表13に登録される障害兆候パターンとしては、例えば、以下に説明するようなものがある。

【0049】

サービスAP3の情報がOS状態情報11のプロセス管理情報を解析しても見つからず、動作記録情報12の警告情報に、サービスAP3の異常終了が記録されているパターン。この場合、サービスAP3の障害と判断する。

【0050】

サービスAP3に対して、一定時間CPU時間の割り当てがなく、サービスAP3の実行が行われていないことが、動作記録情報12に記されたコンテキストスイッチの実行履歴から判るパターン、及び、この場合に、OS状態情報11のプロセス管理情報やロックの使用状況からサービスAP3が何らかの動作の待ち合わせを行っていることが調べられたパターン。これらの場合、サービスAP3に対するCPU時間の割り当てで障害が起きたと判断する。

【0051】

動作記録情報12の割り込みに関する動作の履歴情報から、キーボード24やマウス25、通信装置28からの反応速度が、前状態情報保持部16に記録した定常状態に比べて遅いパターン。この場合、サービス品質の低下や、装置自体の故障の兆候と判断する。

【0052】

動作記録情報12のエラー履歴情報にメモリ21のパリティエラーが報告されているパターン。この場合、メモリ21の故障の兆候と判断する。

【0053】

動作記録情報12にあるプロセス間通信の記録や、第1OS1のメモリ割り当て記録、OS状態情報11のメモリ使用状況から、メモリ不足から何らかのAPが異常終了しているパターン。この場合、メモリリークによる障害の兆候や、過負荷による障害の兆候と判断する。

【0054】

また、OS状態情報11からファイルハンドルなどのシステムリソースに余裕がなくなっているのを検知したパターン。この場合、リソース不足になる可能性があり、障害の兆候と判断する。

【0055】

OS状態情報11のOSのCPU時間の消費を調べることにより、第1OS1自体が消費するCPU時間とサービスAP3が消費するCPU時間の割合を調べて、その割合が、前状態情報保持部16に記録されているものに比べて、極端に第1OS1の割合が多いことを検知したパターン。この場合、サービス品質が低下しはじめている可能性があり、障害の兆候と判断する。

【0056】

動作記録情報 12 の割り込み履歴から、割り込みが多発しているのが検知され、OS 状態情報 11 から第 1 OS 1 が消費する CPU 時間が多いことが検知されたパターン、この場合、割り込み処理の負荷が高くなることによって、サービス品質が低下しはじめている可能性があり、障害の兆候と判断する。

【0057】

動作記録情報 12 のエラー発生履歴情報から、ディスクアクセスエラーが発生していることが検知されたパターン。この場合、そのディスクの物理的な寿命に達した、あるいは、欠陥が生じた可能性があり、障害の兆候と判断する。

【0058】

OS 状態情報 11 の温度センサ情報から、前状態情報保持部 16 に記録している定常状態に比べて CPU 23 や計算機 20 の温度が上昇したり、冷却装置 26 のファンの回転数が減少していることが検知されたパターン。この場合、ハードウェア障害を引き起こす可能性があり、障害の兆候と判断する。

【0059】

OS 状態情報 11 や動作記録情報 12、AP 状態情報 17 等のデータは、第 1 OS 1 においても重要度が高く、アクセス頻度が高いデータである。これらがディスク上へスワップアウトされているパターン。この場合、メモリが深刻な不足状態にあると判断する。

【0060】

次に、図 7 に示す対処プログラム表 14 に登録されている対処プログラムの対処動作の例を、図 7 を参照して説明する。

【0061】

サービス AP 3 が異常終了している場合、プログラム番号 0 のプログラムが、サービス AP 3 を再度起動させる (1400)。

【0062】

サービス AP 3 の実行が止まっている、あるいは、サービス AP 3 が他の AP の実行を待っている場合、プログラム番号 1 のプログラムが、これらの AP の優先度を上昇させるように、補助ドライバ 7 に指示する (1401)。

【0063】

ある装置の故障が予測される場合、プログラム番号 2 のプログラムが、その装置の使用を禁止する。また、メモリ 21 の障害の場合、プログラム番号 2 のプログラムが、対応する部分を割り当てないようにマルチ OS 制御部 5 の OS 間メッセージ通信部 9 を用いて指示する。例えば、外部記憶装置 23 が複数ある場合、1 つに不良が発見されたとしても、他の記憶装置が利用可能ならば、計算機 20 全体を止める必要はない。このため、不良の発生した記憶装置のみを利用不可能にして、計算機の動作を続行させる (1402)。

【0064】

メモリ不足やリソース不足につながる障害の兆候を検知した場合、プログラム番号 3 のプログラムが、メモリやリソースを大量に消費している AP を再起動するように、補助ドライバ 7 に指示する (1403)。

【0065】

また、割り込みが多発している場合、プログラム番号 4 のプログラムが、割り込み処理の原因となる装置への動作を止め、割り込み時の動作を変更することを補助ドライバ 7 に指示する (1404)。

【0066】

冷却装置 26 に障害の兆候が現れた場合、プログラム番号 5 のプログラムが、補助ドライバ 7 を用いて、CPU 22 の動作周波数を下げ (1405)、また、プログラム番号 6 のプログラムが、第 1 OS 1 が動作するときに CPU 22 の一時停止命令を実行する等により、発熱を抑制するように指示する (1406)。

【0067】

前述した障害対処動作が実行できない場合、最終的に、プログラム番号 7 のプログラムが、第 1 OS 1 を再起動させる (1407)。

【0068】

前述において、第1OS1あるいはサービスAP3を再起動させた場合、直前に前状態情報保持部16に退避していたAP状態情報17を用いて、サービスAP3の動作を再開させる。

【0069】

以上説明したような方法を用いることにより、本発明の第1の実施形態は、第1OS1に実際の障害が発生する前に、その兆候を検知し、障害を未然に防ぐように第1OS1を動作させることができる。

【0070】

図9は本発明の第2の実施形態による計算機システムのハードウェア構成を示すブロック図である。なお、第2の実施形態におけるハードウェア構成以外の部分の構成及び動作は、第1の実施形態の場合と同一である。

【0071】

図9に示す本発明の第2の実施形態は、図1に示した第1の実施形態に対して、外部から監視を行うための外部端末40を用意し、解析予測AP4が検出した障害の兆候の情報を外部端末40へ送り、外部端末40からこの障害に対する対応操作を行うことを可能とすると共に、自己回復不能な障害が予測された場合に外部に対応を求めることができるようにしたものである。

【0072】

前記の外部端末40と計算機20とはネットワーク41を用いて通信を行うことが可能であり、相互間での情報の授受を行うことができる。ネットワーク41は、第2OS2のみから利用することができる第2通信装置43に接続されている。このようにすることにより、第1OS1が用いている第1通信装置42に障害の兆候が現れていても、第2OS2から安全にネットワーク41を用いることが可能である。また、この第2の実施形態は、第1OS1と第2OS2とが同一の第1通信装置42を共有してコストを削減するように構成することもできるが、この場合、第1通信装置42の障害以外の障害の兆候が現れた場合にのみ対応することになる。

【0073】

前述したような本発明の第2の実施形態によれば、第1OS1に実際の障害が発生する前に、その兆候を検知して、実際に起きうる障害を外部に通知することができ、管理者が外部にいるような場合にも、外部の管理者が適切な対処を行うことができる。

【0074】

図10は本発明の第3の実施形態による計算機システムのハードウェア構成を示すブロック図である。図10に示す第3の実施形態は、前述の第2の実施形態による計算機システムを複数連結して、サービスをフェイルオーバーする構成としたものである。なお、第3の実施形態におけるハードウェア構成以外の部分の構成及び動作は、第1の実施形態の場合と同一である。

【0075】

図10に示す第3の実施形態による計算機システムは、通常時におけるサービスを行う現用系の第1計算機50と第1計算機50の代用が可能な待機系の第2計算機51とにより構成される。そして、第1計算機50と第2計算機51との間は、第1、第2の計算機のそれぞれが持つ第2OS2からしか使用することができない通信装置43によってネットワーク41を介して接続されている。前述の通信装置43から接続されたネットワーク41には、管理用の外部端末40が接続されている。また、第1計算機50と第2計算機51との間では、外部記憶装置23を共有している。

【0076】

前述したように構成される第3の実施形態による計算機システムにおいて、第1計算機50の第1OS1に障害の兆候が検出されたとき、第1計算機50内の解析予測AP4は、障害が起きる前に、第1計算機50内の第2通信装置43を通じて、第2計算機51に第1計算機50に生じた障害の兆候と、その履歴や状態とを知らせる。第2計算機51

は、これらの情報から第1計算機50の障害に備え、サービス引き継ぎのための処理を行う。また、第2計算機51は、これらの情報から自計算機での同様の障害の発生に対応することができる。例えば、第2計算機51は、障害が起きる可能性がある場合、第1計算機50側のOS状態情報11や動作記録情報12、前状態情報保持部16、AP状態情報17を、予め第2計算機51の解析予測AP4に送信しておき、障害につながる動作を解析しておくことにより、第2計算機51の第1OS1で第1計算機50と同じ障害が発生することを未然に防ぐことができる。

【0077】

図11は前述した本発明の第3の実施形態でのサービス引継ぎのための処理動作を説明するフローチャートであり、次に、これについて説明する。

【0078】

(1) 第1計算機50の第1OS1に障害が検出されたとき、第1計算機50内の解析予測AP4は、まず、外部端末40に切り替えの開始を送信し、第1計算機50内のサービスAP3のサービスを再開させるために、AP状態情報17を第2計算機51に送信する(ステップ140、141)。

【0079】

(2) その後、第1計算機50と第2計算機51とは、ネットワークの設定の引き継ぎと、外部記憶装置の引き継ぎとを行い、最後に、第2計算機51に動作を切り替えて、第1計算機50の第1OS1を停止する(ステップ142～145)。

【0080】

前述したサービス引継ぎの処理では、最初に、外部端末40に切り替えの開始を送信するとして説明したが、前述した切り替えの処理が終了した後、第2計算機51の第2OS2から第1計算機50の第1OS1の障害報告を外部端末40に送信するようにしてもよい。

【0081】

引き継ぎのための情報の第2計算機51への送信は、障害の兆候の情報を第2計算機51に送信したときと同様に、第2通信装置43を介して行われる。

【0082】

また、前述した本発明の第3の実施形態において、第1、第2の計算機50、51内の第2通信装置43を第1OSが使用する第1通信装置42と共用して、前述した障害の兆候の情報、引き継ぎのための情報を送信するようにすることもできる。これにより、それぞれの計算機内に通信装置43を別に用意する必要がなくなる。

【0083】

また、前述した本発明の第3の実施形態において、ディスクの状態を引き継ぐ必要がない場合、外部記憶装置23を共有せず、それぞれ個別の外部記憶装置23を持つようにすることも可能である。この場合、図11に示すフローのステップ143の処理の実行を省略することができる。

【0084】

また、前述した本発明の第3の実施形態において、待機系となる第2計算機51における電力消費を抑えるため、障害による引継ぎが行われるまで第2計算機51の電源を切っておくことが可能である。この場合、本発明により第1計算機50で障害が予想されたときに、第2計算機51を起動し、障害が発生する前に引き継ぎ動作を行うことによりサービスの停止時間を最小に抑えることができる。

【0085】

前述したような構成を持つ本発明の第3の実施形態によれば、現用系の計算機における第1OS1に回避不可能な障害が予測された場合にも、業務を行うサービスAP3の業務処理を、待機系の計算機により継続し続けることができる。

【0086】

前述した本発明の第3の実施形態は、計算機として独立した2台の計算を現用系、待機系として使用するとして説明したが、本発明は、計算機として、論理分割制御部を有する

仮想計算機を用い、マルチOS制御部5として、仮想計算機を構築する論理分割制御部を用いるように構成することもできる。

【0087】

前述の論理分割制御部については、OSシリーズ第11巻VM（岡崎世雄・全先実著：共立出版）（以下、文献2という）に仮想計算機の制御部（CP）として紹介されている。この文献2によれば、CPは、仮想計算機への仮想的な物理メモリ割り当て状態を表すシャドウ・テーブルや、仮想CPUのレジスタの状態を表すVMBLOCKを有する。

【0088】

図12は本発明の第4の実施形態による計算機システムのプログラム構成を示すブロック図である。図12に示す第4の実施形態は、マルチOS制御部5に代わって前述した論理分割制御部を利用して構成したものである。図12に示す本発明の第4の実施形態で利用する論理分割制御部60は、論理分割を行って仮想計算機システムを構成する際に通常設けられているもので、前述したシャドウ・テーブルやVMBLOCKを取得することのできるOS状態取得部59及び他のOSの実行を制御するOS実行制御部58を備えると共に、本発明のためにメモリ取得部8とOS間メッセージ通信部9とを備えている。また、第4の実施形態による計算機システムを構成する各プログラムは、メモリ21の中に格納されている。

【0089】

本発明の第4の実施形態による計算機システムは、論理分割制御部60を用いるため、この計算機上で同時に動作するOSは2つ以上存在する。このうち、通常業務を行うサービスAP3が動作するOS群に属するOSを第1OS1とし、解析予測AP4が動作するOS群に属するOSを第2OS2とする。そして、前述の第1OS1と第2OS2とを1つつつセットにしたものを、仮想マルチOSセットとし、任意のn個の仮想マルチOSセットである第1仮想マルチOSセット61～第n仮想マルチOSセット6nを実現する。

【0090】

論理分割制御部60は、各仮想マルチOSセット61～6nに属する第1OS1と第2OS2とを、交互に切り替えながら実行する。このため、解析予測AP4が動作している間、監視の対象となる第1OS1の動作を確実に止めることができる。これにより、解析予測AP4が第1OS1の状態を解析する間、第1OS1が動作して勝手にメモリ上のデータを書き換えることを防止することができる。

【0091】

前述したような構成を有する本発明の第4の実施形態によれば、論理分割制御部60を持った計算機上でもマルチOS環境を実現することができる。

【0092】

図13は本発明の第5の実施形態による計算機システムのプログラム構成を示すブロック図である。図13に示す第5の実施形態は、前述した第4の実施形態における仮想マルチOSセットを複数個まとめて障害対応グループを作ることにより構成した例である。

【0093】

図13に示す第5の実施形態は、複数の障害対応グループ71～7nと図12に示す場合と同様に構成される論理分割制御部60とにより構成され、第1～第nの複数の障害対応グループ71～7nのそれぞれは、2つ以上（図には2つだけ示している）の仮想マルチOSセット61、62～6n1、6n2により構成されている。各仮想マルチOSセットのうち、通常時にサービスを行う第1OS1を持っているものを現用系システム、現用系の障害発生時にサービスを引き継ぐ他方のものを待機系システムとする。いま、第1障害対応グループ71の現用系システムである仮想マルチOSセット61で障害が予測された場合、同一グループの待機系システムである仮想マルチOSセット62へサービスの引継ぎが行われる。サービス引継ぎのための通信には、論理分割制御部60内のOS間メッセージ通信部9が用いられる。

【0094】

前述したような構成を有する本発明の第5の実施形態によれば、物理的な計算機の台数

を増加させることなく、サービスを安定して提供し続けることのできる計算機システムを構築することが可能となる。

【0095】

図14は本発明の第6の実施形態による計算機システムのプログラム構成を示すブロック図である。図14に示す第6の実施形態は、前述した第5の実施形態における1つの仮想マルチOSセット61を、現用系の第1OS611と待機系の第1OS621と、1つの第2OS2及び解析予測AP4とにより構成したものである。

【0096】

すなわち、図14に示す第6の実施形態は、1つの仮想マルチOSセット61を、現用系の第1OS611と待機系の第1OS621とで、1つの第2OS2及び解析予測AP4を共有するように構成される。そして、解析予測AP4が現用系の第1OS611やサービスAP613に障害を予測した場合、通信装置の代わりに、図14には示していない図12の場合と同様に構成される論理分割制御部60内のOS間メッセージ通信部9により待機系の第1OS621及びサービスAP623にサービスの引継ぎを行った後、OS実行制御部58を使用して現用系の第1OS611を止めて、代わりに待機系の第1OS621を実行する。

【0097】

図14に示した本発明の第6の実施形態は、論理分割された仮想計算機システムにより構成されるとして説明したが、論理分割しなくても、メモリ内に2つの第1OSと、1つの第2OSを格納して、図2に示す場合と同様に構成することもできる。

【0098】

図15は本発明の第6の実施形態で解析予測AP4が使用するデータテーブルを示す図である。

【0099】

図14に示した本発明の第6の実施形態は、1つの解析予測AP4に対して、解析対象となる第1OS1が1またはそれ以上存在する場合、解析に用いる障害兆候表13や障害対処表10、対処プログラム表14、情報アドレス変換表15として、対象となるOS毎に別々のものを用いることにより、解析対象の数の多さに対処する。例えば、対象となる第1OSが2つである場合、解析予測AP4は、図15に示すように、それぞれの第1OSに応じた第1データセット80及び第2データセット81を持つことにより、異なる2つの第1OSに対応する。各データセット80、81の内容は、図6～図8により説明したものと同様である。

【0100】

前述したような構成を有する本発明の第6の実施形態によれば、第2OS2の数を抑えることができるため、第2OSや解析予測APが利用するメモリやディスク等のリソースを少なくすることができる。

【0101】

また、第6の実施形態の変形例として、複数の第1OS1と単体の第2OS2とによりシステムを構成し、単体の第2OS2上で複数の解析予測AP4を動かす方法がある。この例によれば、解析対象OSがあまり多くない場合、第2OS2の数を抑えることができる。

【0102】

なお、前述で説明した本発明の第4～第6の実施形態による計算機システムは、そのハードウェア構成としては、第1の実施形態で説明したものと実質的に同一でよく、また、各実施形態で説明した以外の詳細な動作等も、第1の実施形態で説明したものと実質的に同一である。

【0103】

図16は本発明の第1の実施形態による計算機システムの変形例によるメモリの内部に格納されるプログラムの構成を示す図である。

【0104】

本発明の第1の実施形態では、1つのメモリ内にAP、OS等の全てのプログラムを格納するとして説明したが、本発明は、マルチOSを構成する際に、図16に示すように第2OS2と解析予測AP4とを、第1OS1を格納するメモリ21から物理的に隔離した第2メモリ221を設けて格納し、第2OSへの切り替えを、ハードウェア動作制御部205によって行うように構成することもできる。

【図面の簡単な説明】

【0105】

【図1】本発明の第1の実施形態による計算機システムのハードウェア構成を示すブロック図である。

【図2】メモリ21の内部に格納されるプログラムの構成を示す図である。

【図3】メモリ21に格納される各プログラム内に設けられるデータテーブルの構成を示す図である。

【図4】第2OS2が情報アドレス変換表15を作成、更新する際の処理動作を説明するフローチャートである。

【図5】第2OS2上の解析予測AP4の処理動作を制御するフローチャートである。

【図6】障害兆候表13の構成を示す図である。

【図7】対処プログラム表14の構成を示す図である。

【図8】情報アドレス変換表15の構成を示す図である。

【図9】本発明の第2の実施形態による計算機システムのハードウェア構成を示すブロック図である。

【図10】本発明の第3の実施形態による計算機システムのハードウェア構成を示すブロック図である。

【図11】本発明の第3の実施形態でのサービス引継ぎのための処理動作を説明するフローチャートである。

【図12】本発明の第4の実施形態による計算機システムのプログラム構成を示すブロック図である。

【図13】本発明の第5の実施形態による計算機システムのプログラム構成を示すブロック図である。

【図14】本発明の第6の実施形態による計算機システムのプログラム構成を示すブロック図である。

【図15】本発明の第6の実施形態で解析予測AP4が使用するデータテーブルを示す図である。

【図16】本発明の第1の実施形態による計算機システムの変形例によるメモリの内部に格納されるプログラムの構成を示す図である。

【符号の説明】

【0106】

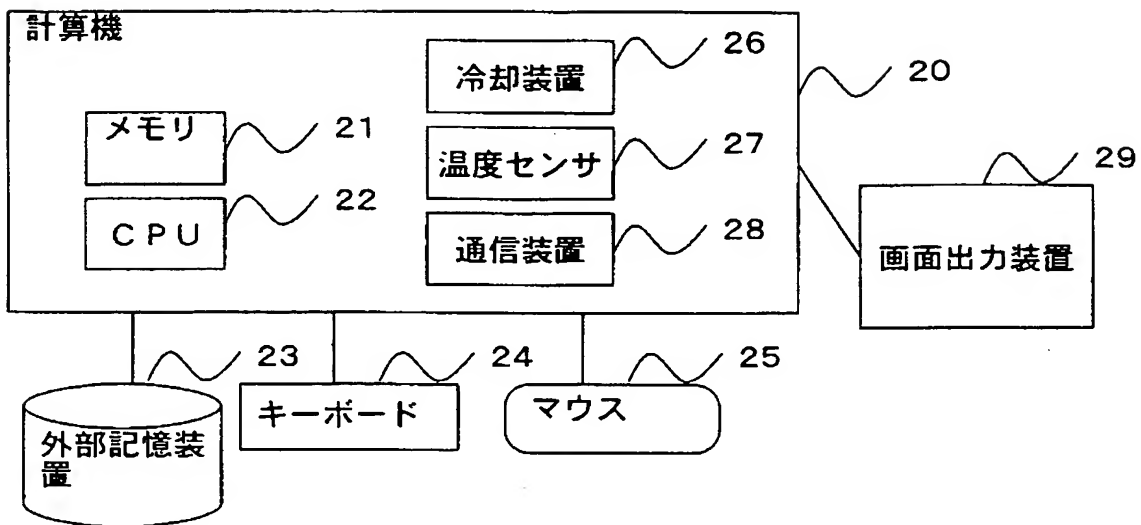
- 1 第1OS
- 2 第2OS
- 3 サービスAP
- 4 解析予測AP
- 5 マルチOS制御部
- 6 動作記録制御部
- 7 補助ドライバ
- 8 メモリ取得部
- 9 OS間メッセージ通信部
- 10 マルチOS制御部インタフェース
- 11 OS状態情報
- 12 動作記録情報
- 13 障害兆候表

- 1 4 対処プログラム表
- 1 5 情報アドレス変換表
- 1 6 前状態情報保持部
- 1 7 A P 状態情報
- 1 8 ページ変換テーブル
- 2 0 計算機
- 2 1 メモリ
- 2 2 C P U
- 2 3 外部記憶装置
- 2 4 キーボード
- 2 5 マウス
- 2 6 冷却装置
- 2 7 温度センサ
- 2 8 通信装置
- 2 9 画面出力装置
- 4 0 外部端末
- 4 1 ネットワーク
- 4 2 第 1 通信装置
- 4 3 第 2 通信装置
- 5 0 第 1 計算機
- 5 1 第 2 計算機
- 6 0 論理分割制御部
- 6 1 第 1 仮想マルチ O S セット
- 6 2 第 2 仮想マルチ O S セット
- 6 n 第 n 仮想マルチ O S セット
- 7 1 第 1 障害対応グループ
- 7 n 第 n 障害対応グループ
- 8 0 第 1 データセット
- 8 1 第 2 データセット
- 2 0 5 ハードウェア動作制御部
- 2 2 1 第 2 メモリ

【書類名】 図面

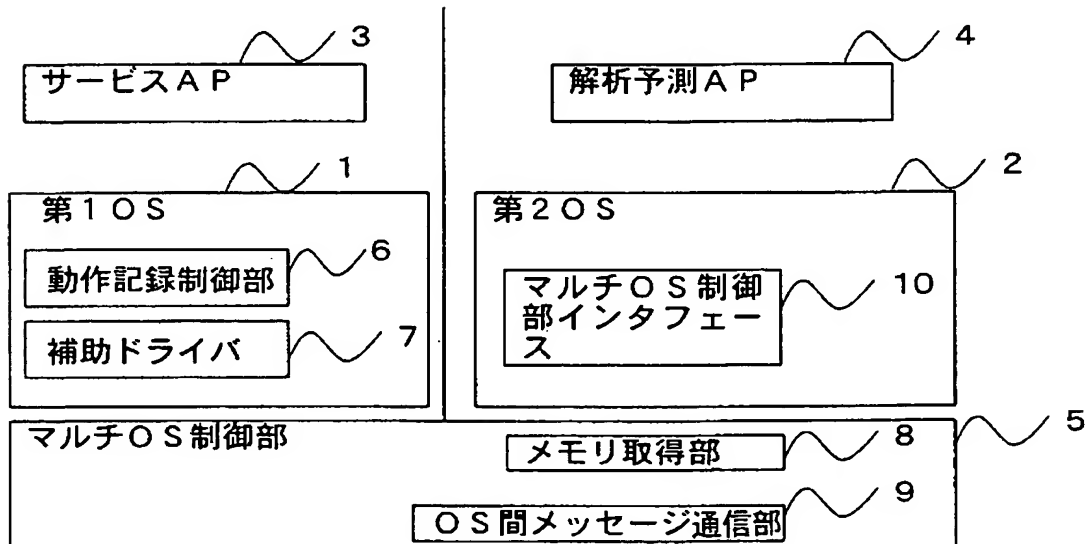
【図 1】

図1



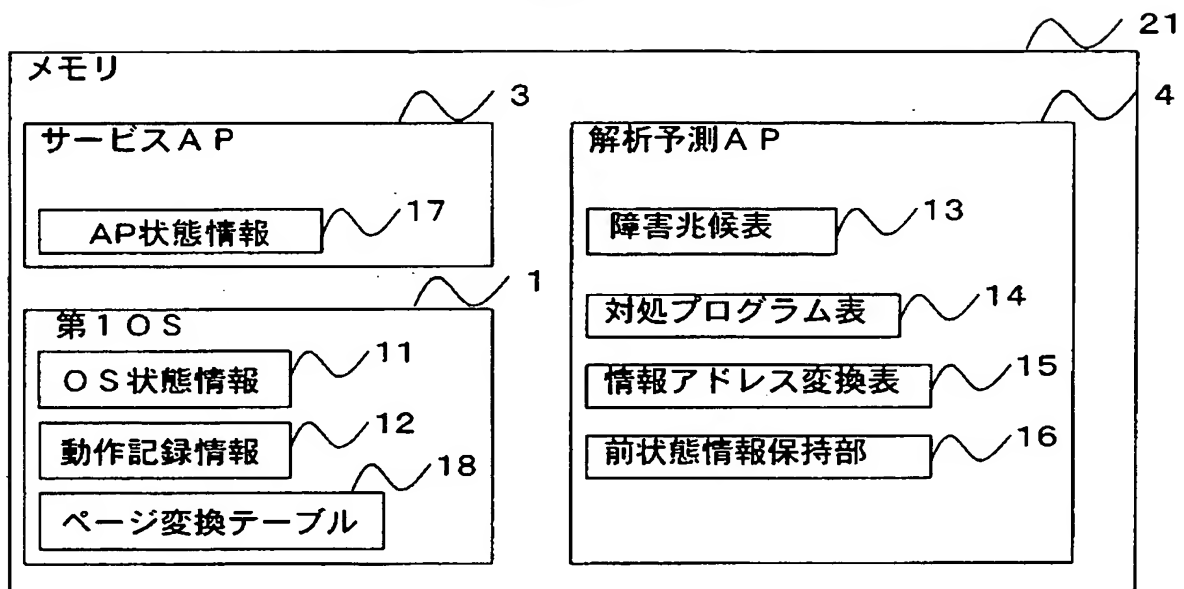
【図 2】

図2



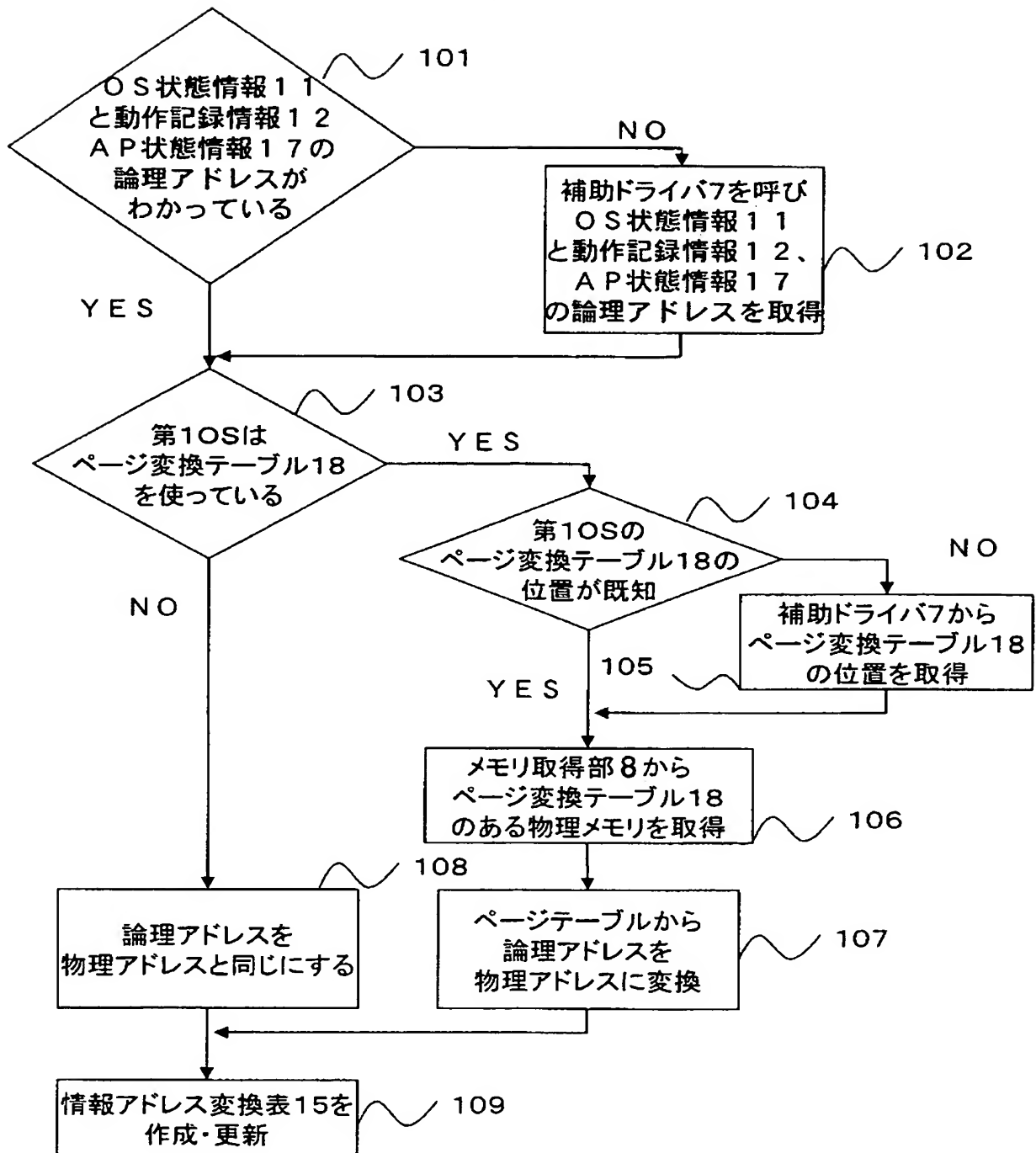
【図 3】

図3



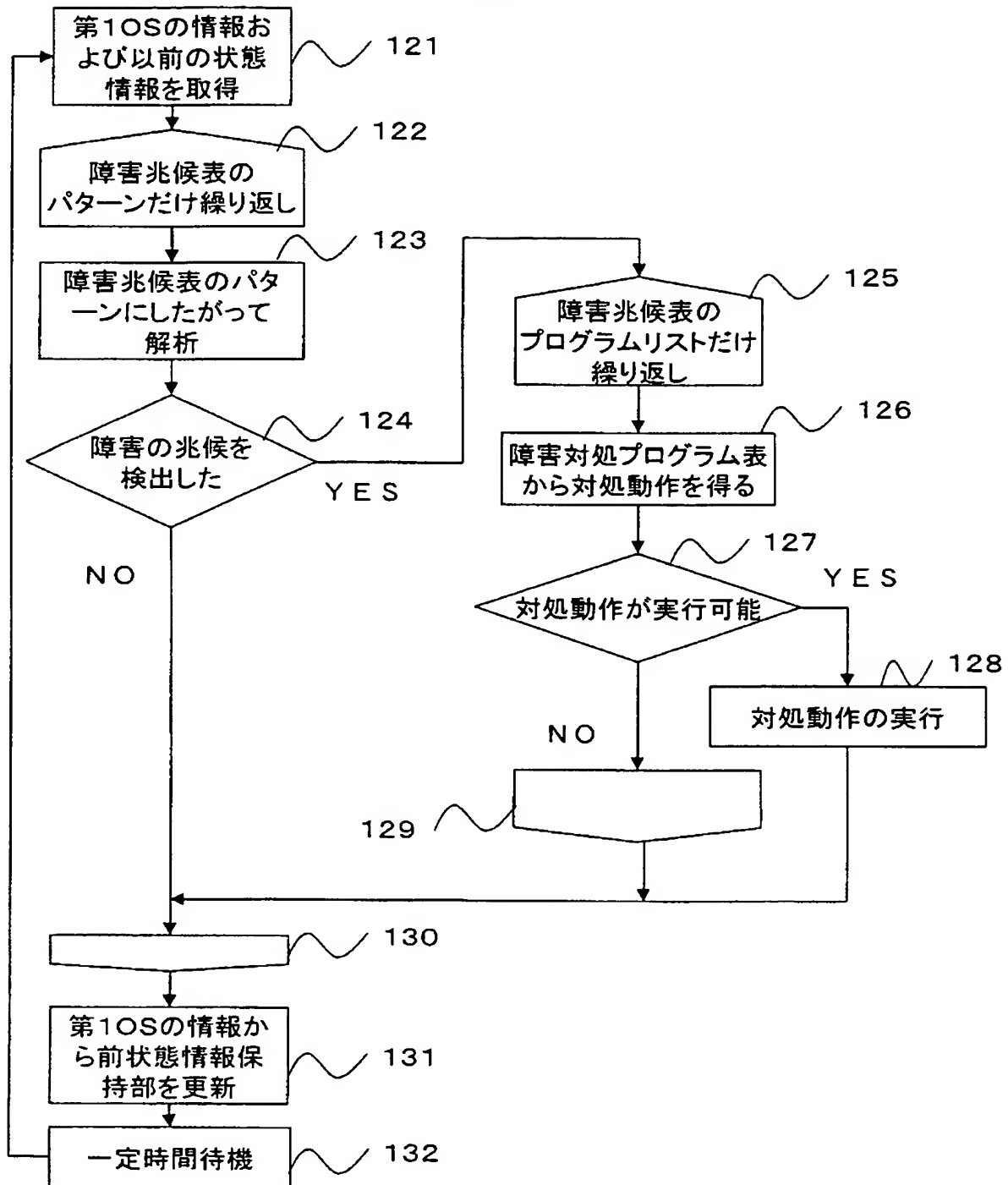
【図 4】

図4



【図 5】

図5



【図 6】

図6

兆候番号	障害の兆候パターン	対処プログラムリスト		
1	動作記録が中断	1	4	
2	デバイスからの異常発生	2	3	
3	割り込みが多発	1	2	4
4	サービスAPの実行が停滞	4		
...		

【図 7】

図7

プログラム番号	対処動作	
0	サービスAPを再起動させる	1400
1	サービスAPの実行優先度を上げる	1401
2	対象装置の使用停止	1402
3	アプリケーション再起動	1403
4	周辺装置へのアクセス禁止	1404
5	CPUの動作周波数を下げる	1405
6	CPUに一時停止命令を実行する	1406
7	第1OSの再起動	1407
...	...	

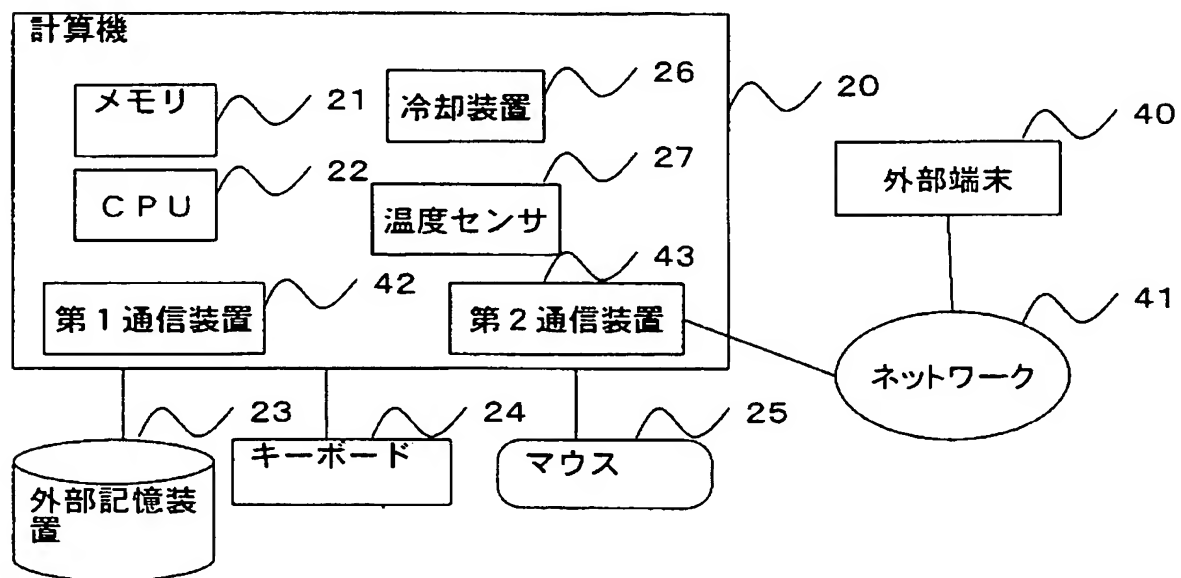
【図 8】

図 8

情報	論理アドレス	物理アドレス
カーネルページディレクトリ	0xC0010000	0x00010000
カーネルページテーブル	0xC0010200	0x00010200
...

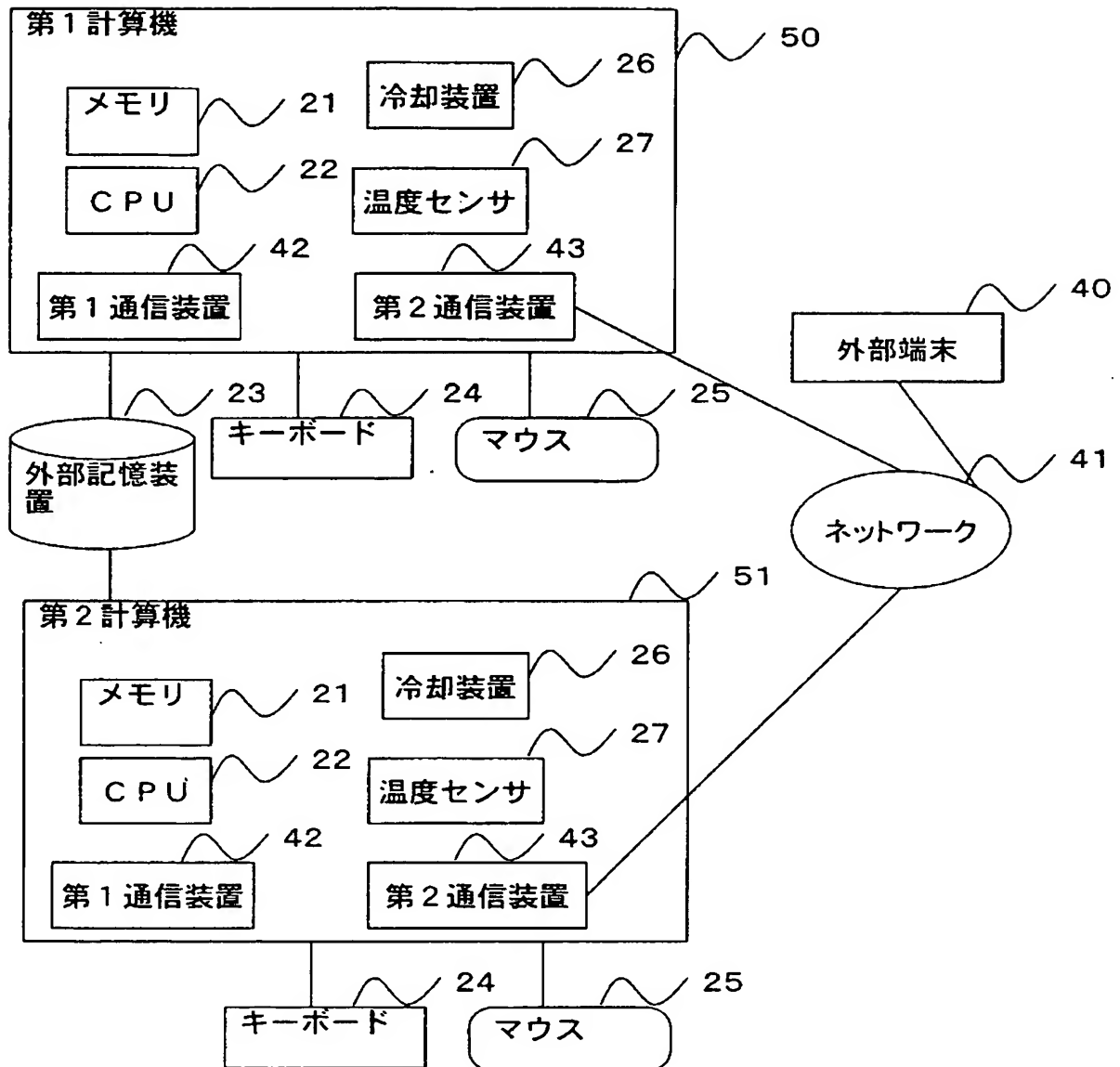
【図 9】

図 9



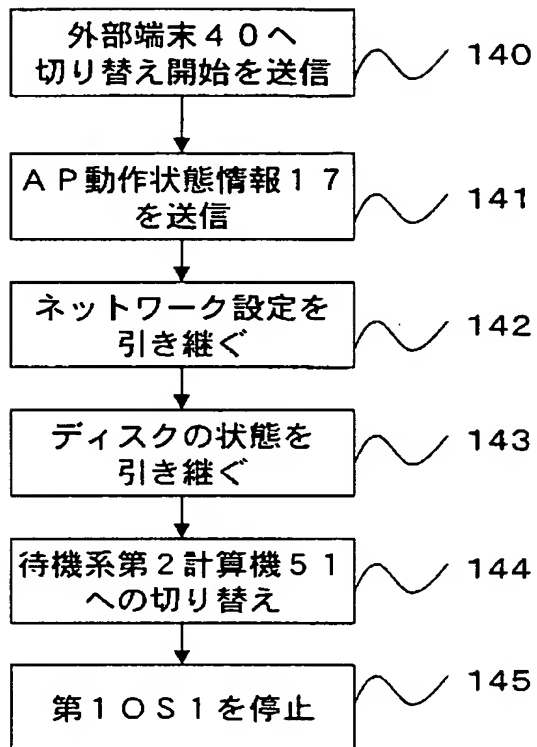
【図10】

図10

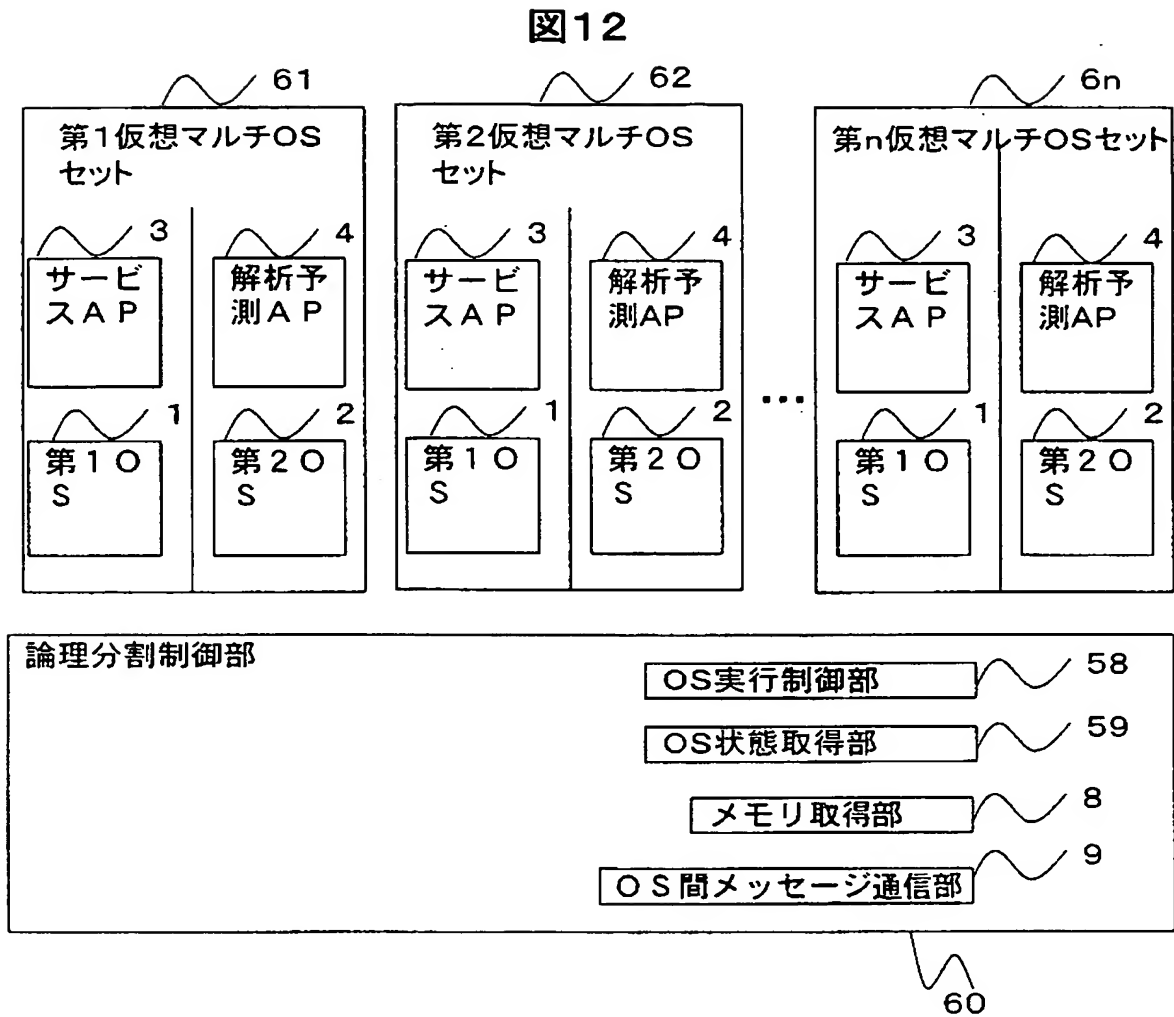


【図 11】

図11

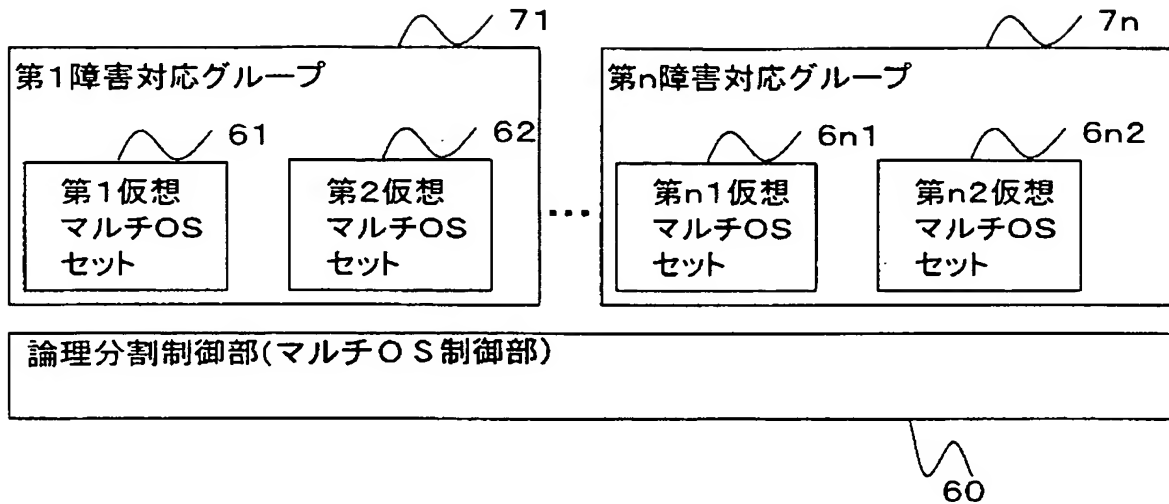


【図 12】



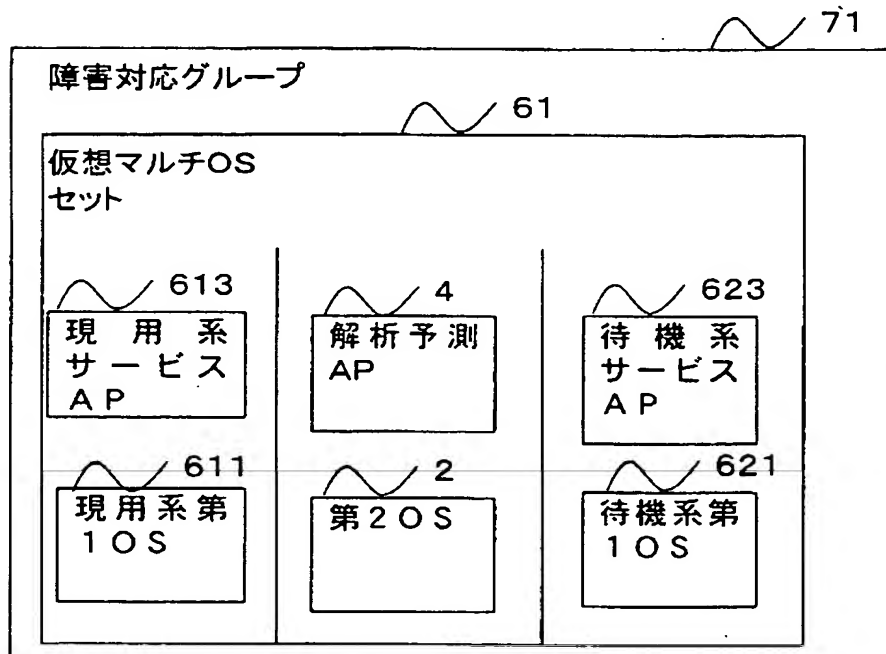
【図 13】

図13



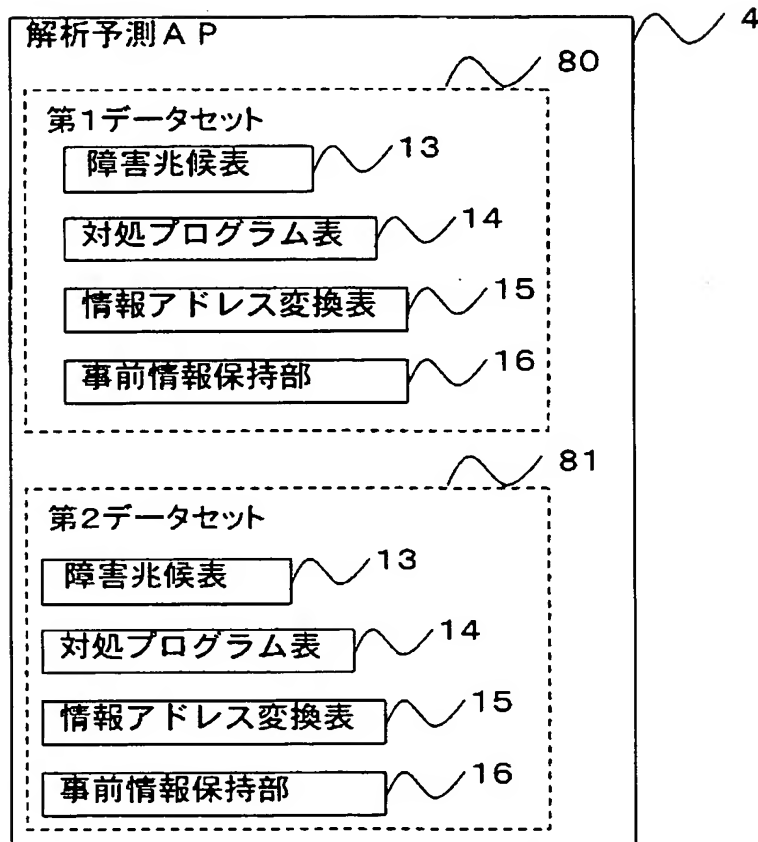
【図 14】

図14

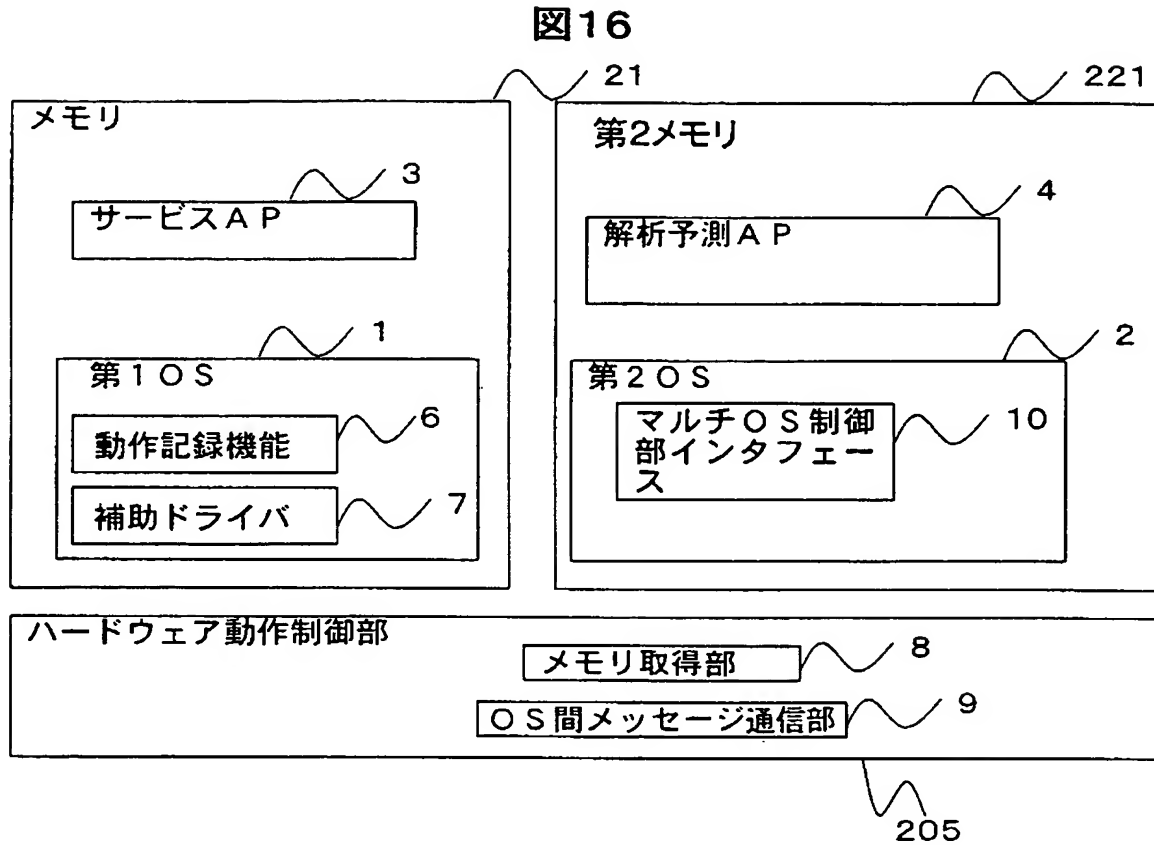


【図 15】

図15



【図 16】



【書類名】 要約書**【要約】**

【課題】動作中のOS及びAPの状態を解析し、障害につながる兆候を検出して、起こりうる障害に対する処理を可能とする。

【解決手段】一般的なOSである第1OS1上で、通常の業務処理を行うサービスAP3が実行される。第1OS1は、第1OS1が行った動作を、動作した時間と共に動作記録情報として記録する動作記録制御部6と、外部からの処理を受け付けて実行する補助ドライバ7とを有する。第1OS1より信頼性の高い第2OS2に備えられるマルチOS制御部インタフェース10は、第2OS2の上で動作するAP4からマルチOS制御部5を操作する。第2OS2上で動作する解析予測AP4は、第1OS1及びサービスAP3の状態を解析して障害の兆候を検知する。障害の兆候を検出した場合、被解析OSやサービスAPの縮退運転や、現用系から待機系の切り替え準備、切り替え等を障害が発生する前に行う。

【選択図】 図2

特願 2 0 0 3 - 3 5 0 8 1 8

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所